

- Consider firm words, such as Corporation or Limited, to be equal to their variations, Corp. or Ltd., during the matching comparison process. To find the abbreviations, the transform uses native script variations of the English alphabets during firm name matching.
- Ignore commonly used optional markers for province, city, district, and so on, in address data comparison.
- Intelligently handles variations in a building marker.

With Japanese data, the Match transform will:

- Block data markers, such as chome and banchi, to be equal to those used with hyphenated data.
- Words with or without Okurigana to be equal in address data.
- Variations of no marker, ga marker, and so on, to be equal.
- Variations of a hyphen or dashed line to be equal.

The Unicode match functionality does not:

- Perform conversions of simplified and traditional Chinese data.
- Compare different scripts, such as Kana to Kanji, or Chinese to English.

The Match transform provides some data normalization options to prepare your data for matching. These options are located in the *Field (Match Input) option* group. Before sending Unicode data into the matching process, you must first separate out the data by country to separate match data flows. The Match Wizard can do this for you when you use the multi-national strategy.

To configure the Match transform for unicode matching 1

1. Use a Case transform to route your data to a Match transform that handles that type of data.
2. Open the *AddressJapan_MatchBatch Match* transform configuration, and save it with a different name.
3. Set the Match engine option in the transform options option group to a value that reflects the type of data being processed. This option is set to Japanese in the *match_starter_unicode* sample. Set any preprocessing options in the Match Criteria Editor. For example:
 - When possible, use criteria for parsed components for address, firm, and name data, such as *Primary_Name* or *Person1_Family_Name1*.
 - If you have parsed address, firm, or name data that does not have a corresponding criterion, use the *Address_Data1-5*, *Firm_Data1-3*, and *Name_Data1-3* criteria.
 - For all other data that does not have a corresponding criteria, use the *Custom* criteria.

Additional Match Features

Typically, matching implies comparing two strings from an alphabetical perspective. In other words, we try to find out how similar two strings are if the strings are compared character by character and calculate a similarity score. But this is not a good approach if the real meaning of the characters is more important. The *semantics* of the data, the real meaning of the data, must also be considered. With proximity matching, the system interprets Geographical, Numeric, and Date type data as such and calculates similarity based on how close they are.